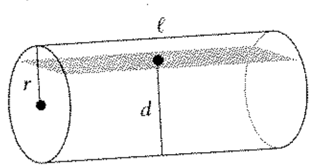
Chapter

3

FUNCTIONS

Functions provide a means of expressing relationships between variables. The values of these variables can be numbers or nonnumerical objects such as geometric figures, functions, or nonmathematical objects. Many of the functions that you have studied in mathematics are *real functions*; that is, functions relating two variables x and y whose values are real numbers. Among the familiar types of real functions are polynomial and rational functions as well as trigonometric, exponential, and logarithmic functions. Typically, real functions are prescribed by formulas of the form $y = f(x)$, such as $y = 2x^2 - x$, $y = \sin(x^2)$, or $y = 2^{-x}$, and much of what you have learned about analyzing and using real functions has depended on making use of formulas such as these.

Figure 1



Even functional relationships that are simple to describe may lead to functional formulas that are relatively complex. Consider, for example, the function that expresses the volume V of fuel in an underground cylindrical tank of length ℓ and radius r whose axis is horizontal, in terms of the depth d of the fuel (Figure 1).

Geometric analysis of this relationship leads to the rather complicated formula:

$$V = f(d) = \ell\left(\pi r^2 - r^2 \cos^{-1}\left(\frac{d}{r} - 1\right) + (d - r)\sqrt{2rd - d^2}\right)$$

Functional relationships of this complexity are often difficult to analyze purely on the basis of formula manipulation. For example, it seems reasonable that the depth d of fuel in the tank is also a function of the volume V of fuel; that is, that the function $V = f(d)$ can be "inverted" to obtain a function $d = f^{-1}(V)$. From its definition, the inverse f^{-1} exists because, for any given volume V of fuel between 0 and the capacity of the tank, pouring that amount of fuel into the tank would fill the tank to one and only one level $d = f^{-1}(V)$. However, solving the equation

$$V = \ell\left(\pi r^2 - r^2 \cos^{-1}\left(\frac{d}{r} - 1\right) + (d - r)\sqrt{2rd - d^2}\right)$$

for d in terms of V is not feasible. Consequently, we cannot find a formula for f^{-1}.

Without a formula, how can we show mathematically that this inverse function exists? The answers to this question and others like it often rest on qualitative features of the functional relationship that may not be apparent from the functional formula. Our look at some familiar real functions in this chapter focuses on features of these functions that may not have been emphasized in your previous courses.

In prior courses, you have also studied many other types of functions that are not real functions. Geometric transformations such as reflections, translations, and rotations are functions relating variables whose values are points or geometric figures. Operations such as addition and multiplication for numbers and functions, the dot product and length for vectors, and determinants of square matrices are functions whose independent and/or dependent variables have values that are not real numbers. In Chapter 2, we used the special type of function called a one-to-one correspondence to show that certain pairs of infinite sets have the same cardinality. Thus, the function concept arises in a wide variety of mathematical contexts. Part of the purpose of this chapter is to illustrate this diversity and to discuss the common ideas about function that are relevant in all or most of these contexts.

Unit 3.1 The Definitions, Historical Evolution, and Basic Machinery of Functions

We begin this chapter by asking "What is a function?" In mathematics, it is common to answer the question "What is a _____?" by giving a definition of _____. We use definitions in mathematics as we do in everyday speech, to help us understand the meaning of an idea and the context in which we use it. Definitions in mathematics also serve another purpose. They are powerful resources for facilitating a proof or solving a problem—providing specific information about the meaning of each condition in the proof or problem.

As important as definitions are, however, they can sometimes mask alternative ways in which an idea can be approached or described, and may not signal why the idea is important. A full answer to "What is a _____?" includes not only a formal definition, but also *alternate definitions and descriptions, why* the idea has been defined, *how* the idea is described, and *to what purposes* the idea is put. A full answer leads us to concept analysis.

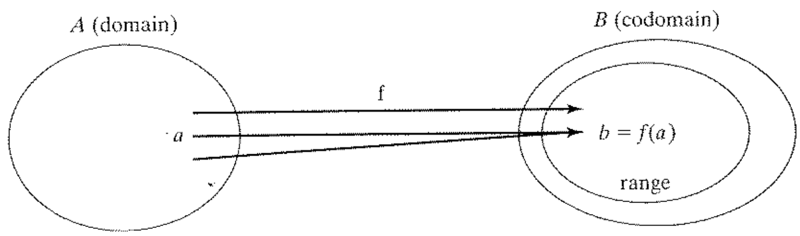
3.1.1 What is a function?

The idea of a function f is to express a relationship between the elements of two sets. If A and B are sets, then a *function f from A to B* is often described as a *rule* or *process* that associates with each element a of the set A one and only one element b of the set B. We often think of each element in the first set as *determining* the corresponding element of the second set.

Definition A function is a rule that assigns to each element of a set A a unique element of a set B (where B may or may not equal A).

The set A is called the **domain** of the function f, the set B the **codomain**, and the subset of the second set B consisting of those elements that are images under the function f of some element of its domain is called the **range** of the function f. When f associates a in A to b in B, the element b is called the **image of a under f** or **the value of f at a**, and a is called the **preimage of b under f**. Figure 2 depicts these sets and elements schematically.

Figure 2



A number of different notations have evolved for functions in mathematics and, more recently, in computer science. Two notations are particularly common in mathematics. When the function f associates a with b, then we write

	f(a) = b,	called **f(x)** or **f() notation,**
or	**f: a → b,**	called the **arrow** or **mapping** notation.

The arrow notation conveys the idea of an action associating the elements from set A to their corresponding values in set B. Some writers use $f\colon A \to B$ only for indicating the domain and codomain, and use $f\colon a \mapsto b$ to identify corresponding elements. We use $\to$ for both. When arrow notation is used, we often say that the function f **maps** the element a onto b and we call f a **mapping** or **map** from A into B. We say that f maps the set A **onto** the codomain B if every element in B is in the range.

A value in the domain of a function is called an **argument** of the function. The variable that stands for the argument is called the **independent variable**. The variable that stands for the values of the function is called the **dependent variable**. In some applications these are called the **input variable** and **output variable**, respectively, reflecting the influence of computer science.

For the function $f\colon x \to y$, many books in precollege mathematics consistently use the single letter f to name the function and distinguish this from the symbol $f(x)$ used to identify the values of the function. In mathematics more broadly, and in computer science, this distinction is not always made, and the symbol $f(x)$ may stand for a function and also its values. Using the symbol $f(x)$ to stand for a function allows the independent variable to be explicitly identified.

A function relationship may be expressed by a *formula* such as $y = x^2$, which squares each real number x, or $y = \sin(x)$, which associates with each real number x the real number y that is the sine of an angle of x radians. A formula that represents a function equates the dependent variable with an algebraic expression written in terms of the independent variable.

A function relationship may also be expressed as a *verbal description* of the correspondence between two sets, such as the correspondence between the set **N** of natural numbers and the set P of prime numbers that associates with the natural number n the nth prime number p_n. In this case, there is no formula for computing the nth prime number p_n for a given natural number n, but the precise meaning of the relationship is nonetheless clear from the description or from the following correspondence diagram.

$$\begin{array}{ccccccccc} \mathbf{N} & 1 & 2 & 3 & 4 & 5 & \cdot\ \cdot & n & \cdot\ \cdot \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \cdot\ \cdot & \downarrow & \cdot\ \cdot \\ P & 2 & 3 & 5 & 7 & 11 & \cdot\ \cdot & p_n & \cdot\ \cdot \end{array}$$

A function relationship may also be expressed by a *table* listing all of its values. For example, the population given by the U.S. Bureau of the Census is a function from the set of years divisible by 10, from 1790 to the present, to the set of natural numbers. This function is partly described in Table 1.

Table 1 U.S. Population $P(t)$ (in Millions) for the Period 1780–1870

Year t	1780[a]	1790	1800	1810	1820	1830	1840	1850	1860	1870
P(t)	2.8	3.9	5.3	7.2	9.6	12.9	17.1	23.2	31.4	39.8

[a] An official census was first taken in 1790, but the Census Bureau has estimated the population for 1780 and earlier.

To allow correspondences where the idea of a function "rule" is not particularly appropriate, mathematicians use a formal definition of function in terms of the language of sets. In this definition, explicitly stated below, we think of each pair of corresponding elements in the two sets as being the components a and b of an ordered pair (a, b) and we think of the function as the set of ordered pairs.

Recall that the **Cartesian product** of two sets A and B, denoted $A \times B$, is the set of *all* ordered pairs whose first components are from A and whose second components are from B. For instance, if $S = \{1, 2\}$ and $T = \{3, 4, 5\}$, then $S \times T = \{(1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5)\}$. The word "Cartesian" in the name of this operation comes from the fact that ordered pairs were first introduced in coordinate graphs on the Cartesian plane. The word "product" and the symbol "$\times$" are quite appropriate for this operation because if A has m elements and B has n elements, then $A \times B$ has mn elements.

Definition

For any sets A and B, a **function f from A to B, f: A → B**, is a subset f of the Cartesian product $A \times B$ such that every $a \in A$ appears once and only once as the first element of an ordered pair (a, b) in f.

For example, let A = the set of all circles in a plane M, and B = the set of all points in M. Then $A \times B$ is the set of all ordered pairs (C, P), where C is a circle and P is a point in M. The subset of $A \times B$,

$$f = \{(C, P) \in A \times B: P \text{ is the center of } C\},$$

is a function $f: A \to B$ because each circle in A has exactly one center. That is, no C is associated with two or more values of P.

On the other hand, the subset of $B \times A$,

$$g = \{(P, C) \in B \times A: P \text{ is the center of } C\},$$

is not a function $g: B \to A$ because any point P is the center of (infinitely) many different circles in the plane. So every point P appears with infinitely many different values of C. However, if we restrict g to the set

$$g_1 = \{(P, C) \in B \times A: P \text{ is the center of } C \text{ and } C \text{ has radius } 1\},$$

then $g_1: B \to A$ is a function.

The ordered pair characterization of function is particularly appropriate for real functions because we can picture the ordered pairs in a *graph*. For this reason, some authors prefer to define a function as a correspondence and define the graph of a function to be the set of ordered pairs created by the correspondence.

The ordered pair definition of function has the advantage that it is precise and unambiguous, but it presents a functional relationship in a rather passive, static way. The description of a function by a rule of correspondence has the advantage that it suggests that a function provides an active procedure for producing range elements from elements of the domain of the function. Together, these descriptions of a function present the concept in a very general yet very precise and useful manner.

Functions and equations

We have noted that some people distinguish $f(x)$ from f, while others do not. It is also the case that some people identify a function with its equation, as in "the function $f(x) = 2x + 5$," while others distinguish a function from a description of its formula or rule. Here we provide a short discussion of notation used in writing functions and equations to clarify the conceptual relationship between functions and equations.

Consider the problem of Jane's average from Chapter 1. Jane has an average of 87 out of 100 after four tests, and we wish to know what score is needed on the 5th test for her average on all 5 tests to be y. (We use y here instead of A.) If Jane scores x points on the 5th test, her average after 5 tests can be written as

(1) $$\frac{4 \cdot 87 + x}{5}.$$

(1) is called *an expression in* x. It is common to use the term "expression" for forms such as (1) that have no equality or inequality signs. Expressions such as (1) are central in that they can be used both in defining a function and in stating an equation, as we now illustrate.

Jane's average after 5 tests is a *function of* the point score on the fifth test. Let us use the symbol f to refer to the function relating x and y. In talking about such a function it is convenient to be able to refer to three different things, the input (independent) variable, the output (dependent) variable, and the function itself. Here is an explicit description of the function f in function notation:

(2) $$f(x) = \frac{4 \cdot 87 + x}{5}.$$

This establishes x as the input variable and f as the name of the function, and links them through the expression in x given in (1).

What we have written in (2) is called a *defining formula* or a *defining equation* for the function f. Perhaps *formula* is the more appropriate word, since *equation* is used universally in a rather different role [see, for example, equation (7), below]. Moreover, an alternate way (3) of giving an explicit definition of the function f uses *mapping notation* and no equation.

(3) $$f: x \rightarrow \frac{4 \cdot 87 + x}{5}$$

This gives the name f to the mapping that sends x to the defining expression (1). Mapping notation also makes clear that there is no notion of equality involved in a function, but only the notion of a relationship between input and output.

Another way of giving an explicit definition of the function f is by relating the expression in x given in (1) to the output variable y:

(4) $$y = \frac{4 \cdot 87 + x}{5}.$$

This establishes y as the output variable and x as the input variable, and links them through the defining expression given in (1). But notice that attempting to link all three of x, y, and f at once in a single formula leads to

(5) $$y = f(x).$$

This is the generic form of function notation and does not define a particular function. To link this to a specific defining expression such as (1) requires a double equality such as

$$y = f(x) = \frac{4 \cdot 87 + x}{5}. \tag{6}$$

Experienced users of mathematics get accustomed to idiosyncrasies of notations for defining functions, but they can create confusion for beginners.

Recall from Chapter 1 that the original question asked what score x Jane needed to average 90. This leads to an *equation in the unknown* x.

$$\frac{4 \cdot 87 + x}{5} = 90 \tag{7}$$

There is a definite connection between the function defined in (6) and the equation stated in (7). The equation amounts to a condition on the input x, namely that the function (6) have the specific numerical output 90. Writing the condition as $f(x) = 90$ would leave out the specific content of the equation. To include this content and also make specific the fact that this is a condition on inputs of the function f, a double equality such as (8) could be used.

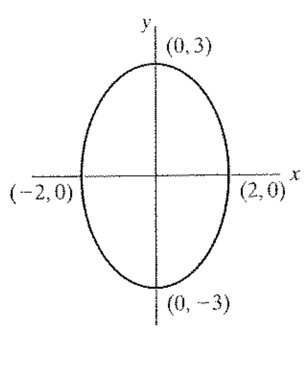
$$f(x) = \frac{4 \cdot 87 + x}{5} = 90 \tag{8}$$

The variable y is commonly used when we discuss the graph of the function f. We can regard (4) above as an alternative way of defining a function, one that makes explicit the name of the output variable y rather than the name of the function. Alternatively, (4) could be regarded as an equation relating the variables x and y. Then the equation (4) also describes the graph of the function as a geometrical object.

Functions of two variables

Equations with two variables (usually x and y) are often used to define lines or curves in the plane, especially when we are interested in them as geometric objects rather than as graphs of functions. In their general form they are not restricted to a $y = \ldots$ formulation. For instance, (9) is an equation that defines the ellipse of Figure 3.

Figure 3

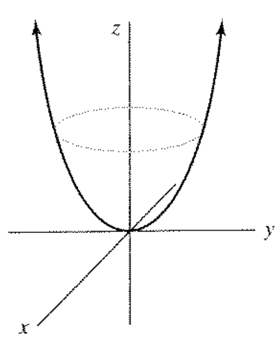


$$\frac{x^2}{4} + \frac{y^2}{9} = 1 \tag{9}$$

Study of equations such as (9), and in general of equations of the form $F(x, y) = k$ (where k is a fixed number), are the subject of the analytic geometry of the plane. They do not necessarily describe a function in the single variable x, because [as is the case in (9)], for some values of x there may exist two values of y. In such cases, the equation $F(x, y) = k$ may identify more than one function of a single variable. For example, equation (9) can be viewed as identifying the union of the two functions f_1 and f_2 defined by

$$f_1(x) = 3\sqrt{1 - \frac{x^2}{4}} \quad \text{and} \quad f_2(x) = -3\sqrt{1 - \frac{x^2}{4}}.$$

Figure 4



Equations of the form $F(x, y) = k$ also serve to identify functions of *two* variables x and y. Specifically, equation (9) is related to the function F of two variables,

$$z = F(x, y) = \frac{x^2}{4} + \frac{y^2}{9}, \tag{10}$$

or, in mapping notation, $F\colon (x, y) \to \frac{x^2}{4} + \frac{y^2}{9}$. The graph of (10) is an elliptic paraboloid above the xy-plane (Figure 4). Equation (9) amounts to a requirement on the output of the function (10), and the pairs (x, y) that meet this requirement are the solutions of (10). To express (9) graphically we can add the plane $z = 1$ to the graph of (10).

down to the xy-plane, is the ellipse of Figure 3.

Sequences

A **sequence** can be defined formally as a function whose domain is the set of integers greater than or equal to a fixed integer k, where k is usually 0 or 1. The image of an integer n in a sequence S is usually denoted by $\mathbf{S_n}$ [rather than $S(n)$] and called the **nth term** of the sequence. The sequence itself is often denoted $\{\mathbf{S_n}\}$ or $\mathbf{S}$. Although a sequence is formally a set of ordered pairs, we often list only the terms in order of n. For instance, the sequence $\{(1,2),(2,4),(3,8),(4,16),\ldots\}$ with rule $S_n = 2^n$ is usually written as $2, 4, 8, 16, \ldots 2^n, \ldots$.

Since sequences are functions, they may be described in the same ways as functions are, by formulas, tables, or correspondences (the correspondence earlier in this section between $\mathbf{N}$ and the set of primes is a sequence). However, sequences possess a fundamental property that distinguishes them from other types of functions—the possibility of being defined *recursively*. For example, the geometric sequence

$$g = \{g_n\} = 2, 6, 18, 54, \ldots, 2(3^n), \ldots$$

has initial term $g_0 = 2$ and common ratio $r = 3$, so that $g_{n+1} = 3g_n$ for all integers $n \geq 1$. These conditions define the sequence *recursively*, that is, after some given terms, later terms are found from earlier terms. The same sequence can also be *explicitly* described by the formula

$$g_n = 2(3^n) \qquad \text{for all integers } n \geq 0.$$

The famous Fibonacci sequence $F = \{F_n\}$ has as its first few terms

$$F = \{F_n\} = 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$$

and can be described recursively by

$$F_1 = 1, \quad F_2 = 1, \quad F_{n+2} = F_{n+1} + F_n \qquad \text{for any integer } n \geq 1.$$

The possibility of defining sequences recursively greatly enhances the utility of sequences for modeling and analyzing problems. Recursive definitions are rooted in the most basic property of the natural number system $\mathbf{N}$—mathematical induction. With mathematical induction, recursively defined sequences become a powerful tool for mathematical modeling and analysis. We examine mathematical induction in some detail in Chapter 5.

A brief history of the concept of function

The function concept was not discovered or conceived by a single individual or at a particular time. Rather, it evolved over a period of several centuries and continues to evolve today in response to important problems in a number of different fields both within and outside of mathematics.

It is interesting that, although the concept and notation for functions were not introduced until the 18th century, graphs of functions were used to analyze their properties as early as the 14th century. An early instance of this use of graphs was by Nicole Oresme (1323–1382), who used velocity-time graphs to study the motions of bodies under uniform acceleration. Although he did not state the law of falling bodies [distance = constant(time)^2] later attributed to Galileo, his results essentially yielded that conclusion.

The historical evolution of the definition of function began with a much less general and less precise description than either the rule of correspondence or the ordered pair definition for a function given earlier in this section. Gottfried Leibniz used the term "function" for the first time in 1694 to describe six very specific line segments associated with a variable point on a given plane curve. (See Project 1 for details about these "Leibniz segments".)

In 1718, Jean Bernoulli (1667–1748) significantly broadened the meaning of function by stating "a quantity composed in any manner whatever of a variable and any constants" is a function of the variable magnitude.[1] He also began experimenting with various notations for functions, with his symbol fx being the closest to the modern $f(x)$ notation.

The notation $f(x)$ for a function of a variable quantity x was introduced in 1748 by Leonhard Euler in his text *Algebra*, which was the forerunner of today's algebra texts. Many other mathematical symbols in use today, including e for the base of the natural logarithm and π for the ratio of the circumference of a circle to its diameter, were introduced by Euler in his writings.

The evolution of the function concept during the eighteenth century took place completely within the context of real functions and centered around a lively interaction between Jean Bernoulli, his sons Daniel (1700–1782) and Nicolaus (1695–1726), Euler, and Jean Le Rond d'Alembert (1717–1783). This interaction was prompted by their common efforts to analyze and describe the motions of a vibrating, tightly stretched string such as a guitar or violin string. As their work progressed, the concept of a function evolved from a rule expressed by a single formula with a finite number of terms, to formulas that would allow infinite series and limits, and finally to piecewise-defined functions that required more than one formula to describe the function. This extension of the meaning of function was used independently in the work of J. B. J. Fourier (1768–1830), whose treatise on heat conduction published in 1822 used infinite series composed of sine and cosine functions (later called *Fourier series*) to represent functions. He observed that his functions included the piecewise-defined functions of earlier mathematicians.

In 1837, Lejeune Dirichlet (pronounced Direesh'lay) (1805–1859) gave the following definition of function: "If a variable y is so related to a variable x that when a numerical value is assigned to x, there is a rule according to which a unique value of y is determined, then y is said to be a function of the independent variable x."[2] Perhaps to emphasize the generality of his definition, Dirichlet introduced the following "badly behaved" function $f\colon \mathbf{R} \to \mathbf{R}$:

$$\begin{aligned} f(x) &= 1 \qquad \text{if } x \text{ is a rational number;} \\ f(x) &= 0 \qquad \text{if } x \text{ is an irrational number.} \end{aligned}$$

This function, which is everywhere discontinuous, now is called the **Dirichlet function**.

Functions commonly studied in calculus may have rather complicated formulas, but they are almost all continuous and also differentiable (except possibly at a finite number of points in their domains). The Dirichlet function is an example of what is sometimes called a *pathological function*. The term "pathological" is used in mathematics for examples that illustrate unexpected or unusual behavior. The Dirichlet function, which is discontinuous at every point, is pathological with respect to continuity. It is perhaps more surprising that there are functions $f\colon \mathbf{R} \to \mathbf{R}$ that are continuous at every point but not differentiable at any point! Of the many such functions, one of the most intuitively appealing is described in Project 5.

---

[1] Cited in Carl B. Boyer, *A History of Mathematics*. Second edition, revised by Uta C. Merzbach. New York: John Wiley, 1991, p. 422.

[2] Boyer, *op. cit*, p. 510.

Until the last half of the 20th century, the concept of function included **multivalued functions**, those in which a domain value may be associated with more than one range value. Examples of rules for multivalued functions $x \rightarrow y$ are $y = \pm\sqrt{x}$, $y =$ a factor of x, and $y =$ an angle whose sine is x. The term "multivalued function" is still common in the study of complex variables, but in most other parts of mathematics these formulas are said to define *relations*, not functions.

The extension of the use of the function concept to contexts in which the domain or range is not necessarily a set of numbers occurred during the latter part of the 19th century and the first half of the 20th century as a result of the developments in set theory, abstract algebra, and analysis. The ordered pair definition of a function $f: A \rightarrow B$ was part of that final evolution to the modern concept. It provided the added generality necessary to make the function concept a central organizational tool and unifying thread for virtually all fields of mathematics.

3.1.1 Problems

1. Show that the formula

$$V = V(d) = \ell\left(\pi r^2 - r^2\cos^{-1}\left(\frac{d-r}{r}\right) + (d-r)\sqrt{2rd - d^2}\right)$$

describes the real function that relates the volume V to the depth d of fuel in an underground cylindrical fuel tank of length ℓ and radius r whose axis is horizontal (Figure 5). [*Hint*: Let $d = BE$, $r = AO = OC$, and $V(d) = \ell$(Area of region $AECBA$), using Figure 6. Notice that Area of region $AECBA = \pi r^2 - 2$(Area of sector ODC) + Area of triangle OAC, and that $\cos(\angle DOC) = \frac{BO}{r}$.]

Figure 5

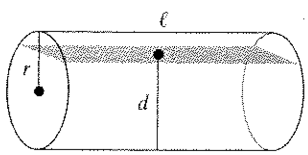
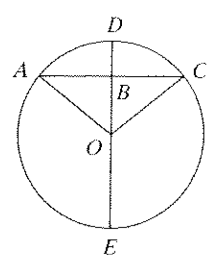


Figure 6



2. The following statements each describe a relationship between two sets A and B in which the given elements are in the set A. Identify the sets A and B precisely in each case. Determine whether or not the relationship defines a function $f: A \rightarrow B$.

Example: The point P is the center of a given circle C in the plane.

Answer: A is the set of circles in the plane. B is the set of points in the plane. The relationship defines a function (because each circle has exactly one center).

a. The triangle T is circumscribed by the given circle C.

b. The circle C circumscribes the given triangle T.

c. The area of a given triangle T is A.

d. The pair $\{p, q\}$ of points in the plane are the foci of a given ellipse E.

3. Give a formula for a function with the indicated domain and range.

a. domain **R**, range **R**

b. domain **Z**, range the set of integers ≥ 2

c. domain **R**, range the set of reals $\geq k$, where k is a given constant

d. domain **R**, range $\{y: a \leq y \leq b\}$

e. domain $\{x \in \mathbf{R}: x > 2\}$, range $\{y \in \mathbf{R}: y > 1\}$

4. a. Suppose A is the empty set. Using either the rule or ordered pair definition of function, are there any functions from A to a set B? If so, characterize them.

b. Suppose B is the empty set. Using either the rule or ordered pair definition of function, are there any functions from set A to set B? If so, characterize them.

5. a. Let A and B be finite sets. Suppose A contains x elements and B contains y elements, with $x \geq y$. How many different functions are there from A to B?

b. Does the answer to part **a** change if $x < y$? Why or why not?

6. Give a precise description of the Euclidean distance d in the coordinate plane as a function of points (x_1, y_1) and (x_2, y_2) in $\mathbf{R}^2$; that is, identify the domain A and range B and a rule for the distance function $d: A \rightarrow B$.

7. a. Give the ordered pairs of the correspondence that maps the letters of the alphabet other than Q and Z onto the telephone digits 2 through 9.

b. Is this correspondence a function? Why or why not?

8. Consider the greatest integer function (or floor function) defined by $\lfloor x \rfloor$ = the largest integer $\leq x$. For example, $\lfloor 1.7 \rfloor = 1$, $\lfloor -\pi \rfloor = -4$. Plot this function with domain the interval $-3 < x < 3$, first by hand and then with a calculator. Explain the difference between the hand plot and the calculator plot.

***9.** The Dirichlet function $f\colon \mathbf{R} \to \mathbf{R}$, which is defined to have the value 1 at all rational numbers and the value 0 at all irrational numbers, can be expressed as a double limit as follows:

$$(*) \quad f(x) = \lim_{m\to\infty}\left(\lim_{n\to\infty}(\cos(m!\pi x))^n\right).$$

Although this expression looks a bit mysterious at first glance, you can unravel the mystery rather quickly as follows:

a. Explain why $(\cos(m!\pi x))^n = 1$ for any integer x and any positive integers m and n with $m > 1$.

b. For any rational number $x = \frac{p}{q}$ with $q > 0$, explain why $(\cos(m!\pi x))^n = 1$ for any positive integers m and n with $m \geq 2q$.

c. For any irrational number x, explain why $-1 < \cos(m!\pi x) < 1$. Then explain why this implies that $\lim_{n\to\infty}(\cos(m!\pi x))^n = 0$ for any positive integer m.

d. Finally, use the results of parts **a–c** together to explain the double limit (*)

3.1.2 Problem analysis: from equations to functions

The purpose of this section is to exemplify once again how pursuing the idea of generalization can raise questions from the level of exercises to the level of mathematical analysis. An earlier example was the problem of matching an average, discussed in Chapter 1. Problems of this sort can be generalized in a systematic way: If the numbers given in the problem statement are replaced by general parameters, the result of the analysis gives the answer as a *function* of these parameters.

A "numbers-in / number-out" problem

We use the following school-level problem to illustrate this idea.

> Person A sets out in a car going at 50 mph. Starting 3 hours later, person B tries to catch up. If person B goes at 75 mph, how long does it take to catch up?

Question 1: Before reading on, solve this problem.

A solution to this problem as stated can be based on the fact that the distances traveled by the two people are the same. Let t be the time it takes person B to catch up. Equating expressions for the two distances leads to a linear equation.

$$(1) \qquad 50\cdot(t+3) = 75\cdot t$$

Solving the equation gives the answer: $t = 6$ hours.

Treated in this way, this is a "numbers-in / number-out" problem. The numbers 50 mph, 3 hours, and 75 mph are the input, and a simple analysis using algebraic techniques produces an output, 6 hours. A particular answer has been given to a particular question.

This type of problem can be useful in giving students practice in setting up and solving equations. Yet if we take the problem one step further to give a general answer to a general problem, much richer mathematics can be illustrated.

A "parameter-in / function-out" problem

As a start on the process of generalizing, suppose we replace person B's speed of 75 mph with a parameter: the speed w.

Question 2: Before reading on, solve the problem in terms of the speed w of person B.

(2) $$50 \cdot (t+3) = w \cdot t$$

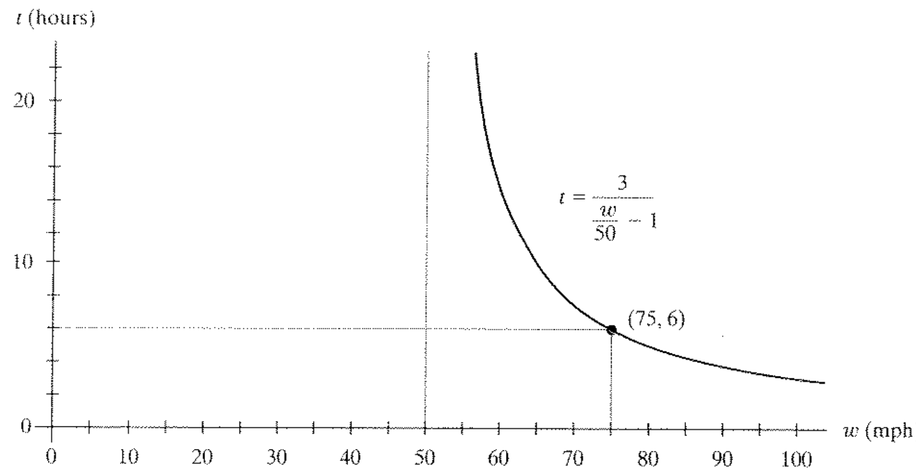
Question 3: Before reading on, solve equation (2) for t.

When we solve equation (2) for t, we do not get a numerical solution. Instead, we obtain $t = \frac{3}{\frac{w}{50} - 1}$, which shows the catch-up time t as a function of any catch-up speed w. The variable w has become the argument of the function

(3) $$t = f(w) = \frac{3}{\frac{w}{50} - 1}.$$

A graph of f can exhibit all solutions. The special case of the original situation (75 mph catch-up speed) and its solution (6 hours) is represented by the single point $(75, 6)$ (see Figure 7).

Figure 7



Having a function as a solution to the problem is much more useful and powerful than having a number as a solution. Formula (3) tells us exactly how the catch-up time depends on the catch-up speed. For example, we can find the slope at $w = 75$ mph (by calculus or estimating from the graph). This slope, about $-\frac{1}{4}\frac{\text{hours}}{\text{mph}}$, tells us that for every one mile per hour person B's catch-up speed increases from 75 mph, about $\frac{1}{4}$ hour is subtracted from the catch up time. We can see also what happens when w is very fast (the time t approaches zero) and what happens when w is just a little over 50 mph (as the catch-up speed w approaches 50 mph, the catch-up time t goes to ∞). Both of these results make sense in the original problem situation.

An "intersection of functions" approach

Another approach to this problem is to model the speed of each car with a linear function. The distance $d_B(t)$ traveled by car B in time t defines a function d_B described by the formula

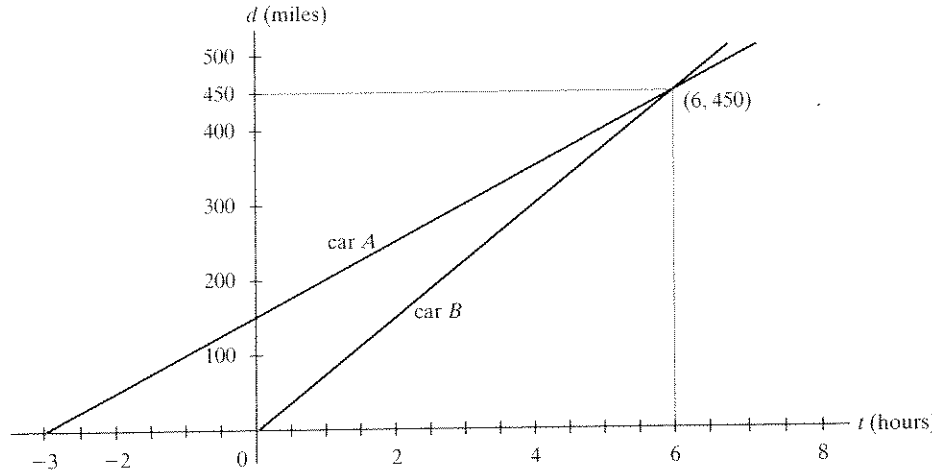
(4) $$d_B(t) = 75t.$$

Similarly, the distance $d_A(t)$ traveled by car A in time t defines a function d_A with

(5) $$d_A(t) = 50(t+3).$$

Figure 8 shows a graph of each of these functions. Notice that negative values of t from -3 to 0 are meaningful, indicating how far car A had traveled t hours *before* car B started.

Figure 8



With this representation, the intersection of the graphs indicates a time when the cars have traveled the same distance. So it is a solution to the original problem. The catch-up time t at the intersection is found by setting $d_B(t) = d_A(t)$, which is equation (1), and which we found to be 6 hours. When $t = 6$, $d = 450$, indicating that both cars have traveled 450 miles.

The graph in Figure 8 generalizes the original problem in a different way than the graph of the function in Figure 7. The vertical distance between the two lines in Figure 8 is the distance between cars A and B at any time t, even when $t > 6$ (assuming the speeds of the cars remain constant).

We have here therefore two different uses of functions and their graphs. A relationship between these two different graphs can be illustrated in a powerful way. Using dynamic geometry software, the graph of Figure 7 can be produced from an animated version of the graph of Figure 8 in which the slope of the function for car B is varied while the t-axis intercept is kept fixed.

Other ways of generalizing the problem

This is not the only way to generalize the problem. To get a fuller picture, note the relevant variables in the problem, as shown in Table 2.

Table 2

Quantity	Symbol	Original Value
speed of person A	v	50 mph
speed of person B (catch-up speed)	w	75 mph
speed increase (how much faster B is than A)	Δv	25 mph
head start time (head start A has over B)	h	3 hours
catch-up time (time for B to catch up with A)	t	6 hours

The function (3) we have derived above gives t as a function of w, while keeping the other variables with their original numerical values (3 hours and 50 mph). Some other ways to generalize are presented in the Problems.

3.1.2 Problems

1. Consider a general "catch-up" situation such as the one analyzed in this section.

a. Show that the time required for a person to catch up is directly proportional to the delay (the elapsed time between the time the first person starts and the time the second person starts).

b. Show that the constant of proportionality in part **a** is dependent only on the ratio of the two speeds.

2. The text claims that, roughly, for every one mile per hour B increases his catch-up speed, the catch-up time is decreased by about $\frac{1}{4}$ hour. What is the basis of this claim? How close is "roughly" if the speed is increased from 75 mph to 80 mph? How close is "roughly" if the speed is decreased from 75 mph to 70 mph?

3. The function f graphed in Figure 7 expresses the catch-up time t as a function of person B's speed w.

a. Find a function that expresses the catch-up *distance* d as a function of person B's speed w. Graph this function, assuming the values $v = 50$ mph for person A's speed and $b = 3$ hours for the delay of person B.

b. Express the catch-up distance d solely in terms of general parameters: person B's speed w, person A's speed v, and the time delay b of person B.

4. The text mentions a way the two graphs in the section could be related using a dynamic geometry program. Carry this out, producing the graph of Figure 7 from a dynamic version of the graph of Figure 8.

5. A general "meeting" problem concerns two people starting off at the same time and heading toward each other.

a. Express the amount of time it will take to meet as a function of the speeds of the people and the initial distance between them.

b. Express the location of the meeting place as a function of the speeds of the people and the initial distance between them.

6. Round trips with and against a wind. Here is a problem of a type most students encounter in their study of elementary algebra.

> An airplane makes a round trip where the one-way distance is 1000 km. On the out-leg the plane faces a headwind of 50 km/h, while on the return there is a tailwind of 50 km/h. If the speed of the plane in still air is 400 km/h, what is the total time for the trip?

a. *A qualitative argument.* Before you solve the problem, think about it in a "qualitative" way: Sketch a rough graph of a function giving the total time for the round trip in terms of the wind speed as the wind speed varies from 0 to 400 kph. Compared with the total time for a round trip with *no* wind, do you think the time for the round trip *with* the wind is (i) less, (ii) the same, or (iii) more?

b. *A numerical answer.* Answer the question of the problem. Does your answer support your qualitative response in part **a**?

c. *A general answer.* The numerical answer does not reveal much about the structure of the situation. Solve the problem again, this time expressing the total time in terms of general parameters for the total distance, the air speed of the plane, and the wind speed. There are many different equivalent symbolic expressions that will express the total time. Try to "coax" the expression you arrive at into a simple form.

d. *The general answer refined.* Express the total trip time with no wind (call it t_0) in terms of the given parameters. Use t_0 to get a more revealing expression for the total time *with* wind. There is a connection of this problem to special relativity through a "Lorenz transformation." Look this up and show what the connection is.

e. *The motion functions.* Functions have not played a role so far in the analysis we have outlined. Give an alternative approach by modeling the situation with the motion functions of the plane's outbound and return trip. Graph these functions.

f. *The dimensionless factor.* A dimensionless factor $\frac{1}{1-r^2}$, where r is the ratio of the wind speed to the plane's speed, appears in the expression for the total time found in parts **c** and **d**. Analyze this factor as a function of r, and graph this function.

7. A mixture problem. Consider this question, also of a common type of problem from elementary algebra.

> How many ounces of a solution that is 90% alcohol need to be mixed with 5 ounces of a solution that is 50% alcohol in order to obtain a solution that is 80% alcohol?

a. Answer the question, letting y be the answer.

b. Generalize the question by replacing 80% by x%. Then graph the function that maps x onto y. Interpret the graph in terms of the original question.

c. Find a formula for the function that maps y onto x and graph that function. Interpret the graph in terms of the original question.

d. Generalize the problem, replacing 90% by A and letting y be the answer. Find a formula for the function that maps A onto y and graph that function. Interpret the graph in terms of the original question.

e. Generalize the problem in a different way, replacing 5 by G and letting y be the answer. Find a formula for the function that maps G onto y and graph that function. Interpret the graph in terms of the original question.

3.1.3 Some types of functions

Functions are important in a wide variety of contexts in which the domains or ranges are not sets of numbers. Functions are also given a variety of other names that are suggestive of the additional special properties that they have. In this section, we explore some of these instances of functions to demonstrate the utility and broad applicability of the concept in mathematics.

One-to-one correspondences as one-to-one functions

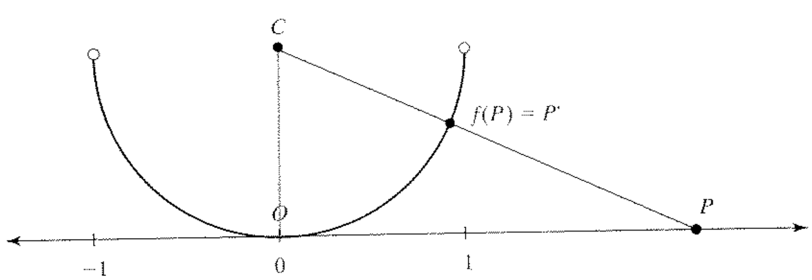
In Chapter 2, we used one-to-one correspondences to establish when sets have the same cardinality. For instance, the existence of a one-to-one correspondence between the set $\mathbf{N}$ of natural numbers and the set $\mathbf{Q}^+$ of positive rational numbers meant that $\mathbf{N}$ and $\mathbf{Q}^+$ have the same cardinality. We indicated that a one-to-one correspondence can be defined by a formula, a diagram, a rule, or a table. One-to-one correspondences are *one-to-one functions*.

Definition A function $f: A \to B$ is a **one-to-one function** or **1-1 function** if and only if every element b in B is the image of *at most* one element a in A. Symbolically, f is 1-1 if and only if for all x_1 and x_2 in A, $f(x_1) = f(x_2)$ implies $x_1 = x_2$.

For instance, with domain $\mathbf{R}$, $y = f(x) = x^3$ defines a 1-1 real function because $x_1^3 = x_2^3$, $x_1 \in \mathbf{R}$, and $x_2 \in \mathbf{R}$ implies $x_1 = x_2$. On the other hand, with domain $\mathbf{R}$, $y = \cos x$ does not define a 1-1 function because $\cos 0 = \cos(2\pi)$ yet $0 \neq 2\pi$.

Often 1-1 correspondences can be presented geometrically. For example, Figure 9 pictures a one-to-one correspondence between the set $\mathbf{R}$ of real numbers and the open interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ in $\mathbf{R}$.

Figure 9



It is described geometrically as follows: Construct a semicircle of radius 1 that is tangent to the real number line at the point O with coordinate 0. Then the line joining a point P on the real line to the center of C of this semicircle determines a point P' on this semicircle. Now $\angle P'CO$ has radian measure between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$, measuring positive and negative from ray CO. Let the coordinate of P correspond to the measure of $\angle P'CO$. For instance, the point with coordinate 1 corresponds to $\frac{\pi}{4}$. This one-to-one correspondence $f: \mathbf{R} \to (-\frac{\pi}{2}, \frac{\pi}{2})$ provides an alternative geometric model of the real number system to the number line—a semicircle—if we identify the point $P' = f(P)$ on the semicircle with the coordinate of P on the number line. In this model, it is natural to interpret the two endpoints of the semicircle as representations of $+\infty$ and $-\infty$. Of course, $+\infty$ and $-\infty$ are not real numbers, but they do represent limits of sets of real numbers. The model has the following property: For any sequence $\{P_n\}$ of real numbers, $\lim_{n\to\infty} P_n = L$ if and only if $\lim_{n\to\infty} f(P_n) = f(L)$. (See Problem 1.)

Inverses of functions

The importance of 1-1 functions lies in the fact that their action can be reversed; that is, given any element y in the range $f(A)$ of a 1-1 function $f\colon A \to B$, there is exactly one element x in A such that $f(x) = y$. The rule that assigns each y in $f(A)$ to the unique x in A for which $y = f(x)$ defines a function with domain $f(A)$ and range A called the *inverse of f*.

Definition If $f = \{(x, y): y = f(x)\}$ and f is one-to-one, then the function $\{(y, x): (x, y) \in f\}$ is called the **inverse** of f and denoted $\mathbf{f^{-1}}$.

More precisely, we have the following:

Theorem 3.1 If $f\colon A \to B$ is a one-to-one function with range $f(A)$, then

$$f^{-1} = \{(y, x) \in f(A) \times A: (x, y) \in f\}$$

is a one-to-one function with domain $f(A)$ and range A.

Proof: To show that f^{-1} is a function from $f(A)$ into A, we need only show that for each y in $f(A)$ there is only one x in A such that (y, x) is in f^{-1}. But

$$(y, x) \in f^{-1} \Leftrightarrow (x, y) \in f$$

and for each y in $f(A)$ there is exactly one x in A such that $(x, y) \in f$ because f is a one-to-one function. By similar reasoning, we see that f^{-1} is a one-to-one function because f is a function.

If $f(x_1) = f(x_2) = y$ and $x_1 \neq x_2$, then if g were the inverse of f, we would have $g(y) = x_1$, and $g(y) = x_2$, and g would not be a function. So a function that is not one-to-one does not have an inverse. ∎

Composition of functions

Simple functions are often combined to produce other functions. *Function composition* gives the result when a second function operates on the images of a first function. Here is a formal definition in the language of ordered pairs.

Definition If $f\colon A \to B$ and $g\colon B \to C$ are functions, then the **composite function** $(g \circ f)\colon A \to C$ is the subset $g \circ f$ of $A \times C$ defined as follows:

$$g \circ f = \{(a, g(f(a))) \in A \times C: a \in A\}.$$

Read "$g \circ f$" as "the composite of f followed by g". We distinguish the operation "composition" from the result "composite" of performing the operation. Some authors use the word "composition" for both.

For example, if $f\colon x \to x^2$ (the squaring function) and $g\colon x \to x - 5$ (the "subtracting 5" function), then $g(f(x)) = g(x^2) = x^2 - 5$, so $g \circ f\colon x \to x^2 - 5$.

Inverse functions and composition of functions

The term "inverse function" comes from the fact that the composite of a function and its inverse is an identity function under the operation of function composition.

The function $f: x \to x^2$ with domain **R** does not have an inverse, but if we restrict its domain to $[0, \infty)$, then the restricted function does have an inverse. In general, if $f: A \to B$ is a function and if C is a subset of A, then the set

$$f_C = \{(a, b): a \in C\}$$

is a function from the set C into B called the **restriction of f to C**. The restriction $f_C: C \to B$ has the equation

$$f_C(x) = f(x) \qquad \text{for all } x \text{ in } C.$$

For any set C, the symbol I_C denotes the **identity function** on C; that is,

$$\mathbf{I}_C = \{(x, x): x \in C\}.$$

Equivalently, I_C is the function from C to C that satisfies $I_C(x) = x$ for all $x \in C$. $\mathbf{I}_C$ maps any element of C onto itself.

With this language, we can connect function composition, inverse functions, and identity functions.

Theorem 3.2 Suppose $f: A \to B$ is a given function. Then there is a function $g: f(A) \to A$ such that

$$g \circ f = I_A \quad \text{and} \quad f \circ g = I_{f(A)}$$

if and only if f is a one-to-one function and $g = f^{-1}$.

Proof:

($\Rightarrow$) Suppose f is 1-1 and $g = f^{-1}$. Then for all $a \in A$, if $f(a) = b$, then $g(b) = a$. So $g(f(a)) = g(b) = a$. Thus $g \circ f = I_A$. Similarly, $f(g(b)) = f(a) = b$ for any element b of $f(A)$, so $f \circ g = I_{f(A)}$.

($\Leftarrow$) This is left to you as Problem 3. ┘

By substituting f^{-1} for g in Theorem 3.2, we see that a 1-1 function f with domain A satisfies the following properties of composition:

$$f^{-1} \circ f = I_A \quad \text{and} \quad f \circ f^{-1} = I_{f(A)}.$$

Thus, just as with the operations of addition and multiplication of real or complex numbers, composing a function with its inverse in either order results in an identity for that operation.

Operations as functions

From the addition of whole numbers that you learned while very young through the operations of differentiation and integration in calculus, you have encountered a host of *operations*. Most of the common operations you have encountered are *unary* or *binary operations*, and both unary and binary operations are functions.

Binary operations are special types of functions of two variables. A **binary operation** takes as input an ordered pair of elements and from them yields a single element as output. If both components of the ordered pair and the output are from the same set S, we say the binary operation is *on the set* S. In function language, **a binary operation on a set S** is any function $S \times S \to S$. (Recall that $S \times S$ is the set of ordered pairs of elements S.) For instance, the operation A of addition can be described in function language as $A: (m, n) \to m + n$. Addition, subtraction, and multiplication are binary operations on the set of integers. Division is not a binary operation on the set of integers, because there are integers whose quotient is not an integer, but division is a binary operation on the set of nonzero rational numbers. (*Note*: Some authors do not require that the output be in S to have a binary operation on S. Instead, they specify that the operation is called **closed** if the output is in S.)

Binary operations need not be on numbers. For example, union and intersection are binary operations on the set of all subsets of a set. Composition of real functions is also a binary operation.

Unary operations are special types of functions of one variable. A **unary operation** takes as input a single element and from it yields another single element. If both the domain and codomain are the same set A, then the unary operation is said to be *on the set A*. That is, a **unary operation on a set A** is any function $u: A \rightarrow A$ that maps the elements of the set into elements of that set. For instance, the reciprocal operation r defined by $r(x) = \frac{1}{x}$ is a unary operation on the set of nonzero rational numbers. On the other hand, r is not a unary operation on the set of nonzero integers because there are nonzero integers whose reciprocals are not nonzero integers.

As with binary operations, unary operations need not be on numbers. For example, given a subset A of a set S, the **complement of A in S**, sometimes denoted by $\overline{\mathbf{A}}$, is the set of elements of S not in A. The function $c: A \rightarrow \overline{A}$, "taking the complement," is a unary operation on the subsets of a set.

Every function $f: \mathbf{R} \rightarrow \mathbf{R}$ is a unary operation on the set of real numbers. For instance, the function f with rule $f(x) = x + 5$ might be described as the unary operation "adding 5." More generally, for any set S, any function $f: S \rightarrow S$ is a unary operation on S.

Operations are often discussed not as special types of functions, but rather with a special terminology of their own. For instance, the idea of *closure* is common with operations, but not with functions. Let f be a unary or binary operation on a set A. If B is a subset of A, then B is said to be **closed with respect to the operation f** if and only if the restriction of f to B is an operation on B.

For example, for the binary operation of addition on the set $\mathbf{Z}$ of integers, the subset E of even integers is closed with respect to addition, but the subset O of odd integers is not. That is, the restriction of the binary operation of addition $A: \mathbf{Z} \times \mathbf{Z} \rightarrow \mathbf{Z}$ to $E \times E$ is a binary operation on E, while its restriction to $O \times O$ is a binary operation but it is not a binary operation on O. Similar conclusions hold for the binary operation of multiplication on the set $\mathbf{R}$ of real numbers on the subset $\mathbf{Q}$ of rational numbers (closed) and the subset $\overline{\mathbf{Q}}$ of irrational numbers (not closed).

Inverse operations are not necessarily inverse functions

Sometimes operations use the same terminology as functions, but with different meanings. Specifically, certain pairs of unary or binary operations are called *inverses* of one another even through they are not pairs consisting of a function and its inverse function. Examples of such pairs are the operations of addition and subtraction for numbers; and the operations of multiplication and division for numbers. In each of these examples, the term "inverse" is used differently for operations than for functions.

For example, the binary operation of addition on the set $\mathbf{Z}$ of integers, the set $\mathbf{Q}$ of rational numbers, and the set $\mathbf{R}$ of rèal numbers does not have an inverse as a function because it is not one-to-one. For instance,

$$7 = 4 + 3 = 8 + (-1) \Rightarrow 7 \text{ is the image of } (4, 3) \text{ and } (8, -1) \text{ under addition.}$$

However, there is a natural way in which addition does have an inverse in the context of functions: For a given set S (such as $\mathbf{Z}$, $\mathbf{Q}$, or $\mathbf{R}$ but not $\mathbf{N}$) on which addition is a binary operation, and a given k in S, define

$$A_k: S \rightarrow S \quad \text{and} \quad S_k: S \rightarrow S$$

by $A_k(x) = x + k$ and $S_k(x) = x - k$ for all x in S. We might call A_k the "adding k" function and S_k the "subtracting k" function. Then A_k and S_k are a pair of inverse functions for each k in S because if $A(x) = x + k = y$, then $S(y) = y - k = x$. You are asked to describe a similar interpretation of multiplication and division in Problem 6.

In calculus, the operations of differentiation and integration are often described as inverse operations as a means of summarizing the statement called the *Fundamental Theorem of Calculus*. We state that theorem here without proof.

Theorem 3.3 **(Fundamental Theorem of Calculus):** If f is a real function that is continuous on an interval $[a, b]$ and if F is the real function defined for each x in $[a, b]$ by

$$F(x) = \int_a^x f(t)\,dt,$$

then $F(x)$ is differentiable on $[a, b]$ and $\frac{d}{dx}\left(\int_a^x f(t)\,dt\right) = f(x)$ for all x in $[a, b]$.

It looks as if integration and differentiation are inverse operations on functions, since integration followed by differentiation yields the original function. That is, $\frac{d}{dx}\left(\int_a^x f(t)\,dt\right) = f(x)$ for any continuous function on $[a, b]$. But since $\int_a^x f'(t)\,dt = f(x) - f(a)$, the value of the composite function in the other order differs by a constant from the "original" function f. In fact, since $\frac{d}{dx}f(x) = \frac{d}{dx}(f(x) + c)$ for any constant c, the unary operation of differentiation does not have an inverse. Thus, to describe integration and differentiation as inverse operations is technically not valid.

Transformations

Some functions that suggest a change of a set, often geometric in nature, are called *transformations*. The most common transformations in school mathematics are *geometric transformations*. A **geometric transformation** is a function whose domain and range are sets of points. Most often the domain and range of a geometric transformation are both $\mathbf{R}^2$ or both $\mathbf{R}^3$. Often geometric transformations are required to be 1-1 functions, so that they have inverses. Examples of geometric transformations are reflections, rotations, size changes, scale changes, and shears. The name "transformation" for these functions comes from the fact that figures in the domain are viewed as having been transformed by the function into their corresponding image figures in the range, and the requirement that they be 1-1 ensures that there is a unique transformation "back" from the image to the original figure.

A rule for a geometric transformation of the plane requires that for each point there be a way to find its image. This rule may be given in geometric language. For example, part of the definition of the reflection r in the plane over line m is that the image of point P is the point P' such that m is the perpendicular bisector of $\overline{PP'}$. Function notation is commonly used, so if r is the reflection over m, we may write $r: P \rightarrow P'$, and $r(P) = P'$. A rule may also be given in terms of coordinates of points. For instance, the transformation T under which the point $(x + 2y, y)$ is the image of (x, y) can be described as $T: (x, y) \rightarrow (x + 2y, y)$, or $T(x, y) = (x + 2y, y)$. This transformation happens to be a shear.

Transformations may be composed. For example, we can compose

$$r_{x=y} = \text{reflection over the line } x = y$$

and $$r_x = \text{reflection over the } x\text{-axis}$$

to obtain $r_{x=y} \circ r_x$, the rotation of 90° counterclockwise about $(0, 0)$, or $r_x \circ r_{x=y}$, the rotation of 90° clockwise about $(0, 0)$. Composition of transformations is basic to the study of congruence and similarity (see Chapters 7 and 8).

Transformations may describe functions in areas of mathematics outside geometry, but almost always from a situation in which there is underlying geometry. For instance, linear transformations are particular functions that map vectors onto vectors, or matrices onto matrices.

Morphisms

Recall from Chapter 1 that $\langle \mathbf{R}^+, \cdot \rangle$ and $\langle \mathbf{R}, + \rangle$ are sets with operations on their elements satisfying certain general rules. The set $\mathbf{R}^+$ of positive real numbers is a group (it satisfies the group properties) under the binary operation of multiplication, while the set $\mathbf{R}$ of real numbers is a group for the binary operation of addition. The function $\log_b: \mathbf{R}^+ \to \mathbf{R}$ is a one-to-one correspondence between $\mathbf{R}^+$ and $\mathbf{R}$ under which results of the operations correspond. That is, if $\log_b(x) = m$ and $\log_b(y) = n$, then the log of the product in $\mathbf{R}^+$ corresponds to the sum of the logs in $\mathbf{R}$: $\log_b(xy) = m + n$. Such one-to-one correspondences between algebraic systems are called **isomorphisms**, and the two systems are said to be **isomorphic** algebraic systems. The term *isomorphism* (from the Greek "iso", meaning "same," and "morph", meaning "form or structure") indicates that the two systems $\langle \mathbf{R}^+, \cdot \rangle$ and $\langle \mathbf{R}, + \rangle$ are algebraically identical. See Problem 15 for another example of isomorphism.

The term **homomorphism** is used for functions from one algebraic system onto another under which results of operations correspond, but that are not necessarily one-to-one functions. For example, if m is an integer greater than 1, the function $h: N \to N_m$ maps any natural number n onto its congruence class n^* modulo m (see Section 6.1.2). Both the set N of all natural numbers and the set N_m of congruence classes modulo m are groups under addition. The function h is operation preserving: $h(a + b) = (a + b)^* = a^* + b^* = h(a) + h(b)$ for all a and b in N.

3.1.3 Problems

1. Consider the correspondence $f: \mathbf{R} \to (-\frac{\pi}{2}, \frac{\pi}{2})$ described in this section.

a. Find the values of $f(-1)$ and $f(10)$.

b. Generalize part **a** to derive a formula for $f(x)$.

c. Use part **b** or a geometric argument based on Figure 9 to explain why $\lim_{n\to\infty} P_n = L$ if and only if $\lim_{n\to\infty} f(P_n) = f(L)$.

2. a. Prove that any two open intervals (a, b) and (c, d) in the set $\mathbf{R}$ of real numbers have the same cardinality.

b. Find a formula for a 1-1 correspondence that demonstrates part **a**.

3. Prove the ($\Leftarrow$) direction of Theorem 3.2, that if $f: A \to B$ is a given function and there is a function $g: f(A) \to A$ such that $g \circ f = I_A$ and $f \circ g = I_{f(A)}$, then f is a one-to-one function and $g = f^{-1}$.

4. Decide if the following subsets of the set $\mathbf{Z}$ of integers are closed for the unary operation of squaring. Support your conclusion.

a. the set E of even integers

b. the set O of odd integers

c. the set P of prime (positive) integers

d. the set C of composite positive integers

5. Describe each of the following as a unary or binary operation on an appropriate set of numbers.

a. the greatest common divisor $\gcd(x, y)$ of x and y

b. the maximum $\max\{x, y\}$ of x and y

6. Provide an interpretation of the statement in the context of functions:

The binary operations multiplication and division of real numbers are inverse operations.

7. Explain in the context of functions the meaning of the following statement:

The reciprocal operation $a \to \frac{1}{a}$ *and the squaring operation* $a \to a^2$ *commute on the set of nonzero rational numbers, but the "taking the opposite" operation* $a \to -a$ *does not commute with the squaring operation.*

8. Let P be a binary operation on the set of positive real numbers such that $P(x, y) = 2x + 2y$.

a. Is P commutative?

b. Is P associative?

c. Give a geometric application of P and interpret the result of part **a** in terms of that application.

9. Repeat Problem 8 for the binary operation A on the set of positive real numbers such that $A(x, y) = xy$.

10. Let S be a finite set with n elements.

a. How many binary operations are there on $S \times S$?

b. How many of these are commutative?

11. The number of permutations of n objects taken r at a time, written nPr or $P(n, r)$, can be considered as a function of two variables n and r described by the rule $P(n, r) = \frac{n!}{(n-r)!}$, where $n \geq r$. P also can be considered as a binary operation.

a. Is P commutative?

b. Is P closed on $\mathbf{N}$?

12. Answer the questions of Problem 11 for the number of combinations of n objects taken r at a time, written nCr, $C(n, r)$, or $\binom{n}{r}$, described by the rule $C(n, r) = \frac{n!}{r!(n-r)!}$, where $n \geq r$.

13. Consider the binary operation of powering (or exponentiation) defined by $p: (x, y) \rightarrow x^y$.

a. Is p a binary operation on $\mathbf{R}^+ \times \mathbf{R}^+$, where $\mathbf{R}^+$ is the set of positive real numbers?

b. Explain why p is not a binary operation on $\mathbf{R} \times \mathbf{R}$.

c. What value(s), if any, does your calculator give for $p\left(-8, \frac{1}{3}\right)$ and $p\left(-8, \frac{2}{6}\right)$?

d. Explain what your answer to part **c** means in the context of binary operations.

14. Let X be a set with 5 distinct elements a, b, c, d, e, and let Y be the set of 5 distinct prime numbers $2, 3, 5, 7, 11$. The set $P(X)$ of all subsets of the set X is an algebraic system for the binary operations of union $\cup$ and intersection $\cap$ of sets. The set $D(2310)$ of all positive divisors of 2310 is an algebraic system for the binary operations of least common multiple (lcm) and greatest common divisor (gcd). (2310 is the product of 2, 3, 5, 7, and 11.)

a. Explain why the sets $P(X)$ and $D(2310)$ both have $2^5 = 32$ elements.

b. Show that the algebraic systems $\langle P(X), \cup, \cap \rangle$ and $\langle D(2310), \text{lcm}, \text{gcd} \rangle$ are isomorphic by defining a one-to-one correspondence f between $P(X)$ and $D(2310)$ such that

$$f(A \cup B) = \text{lcm}(f(A), f(B)) \quad \text{and}$$
$$f(A \cap B) = \text{gcd}(f(A), f(B))$$

for all subsets A and B of X.

15. Let i be the complex number such that $i = \sqrt{-1}$, and let S_4 be the set $\{1, i, -1, -i\}$.

a. Prove that the operation $\cdot$ of multiplication of complex numbers is a binary operation on S_4 and that $\langle S_4, \cdot \rangle$ is a group.

b. Prove that the function $h: Z \rightarrow S_4$, defined by $h(n) = i^n$ for each integer n, is a homomorphism of $\langle N, + \rangle$ onto $\langle S_4, \cdot \rangle$.

Unit 3.2 Properties of Real Functions

We use the term **real function** to refer to a function whose domain and range are subsets of the set $\mathbf{R}$ of real numbers. The function f that opens this chapter and all the functions of the problem analysis of Section 3.1.2 are real functions. A function may involve only real numbers but still not be a real function. For instance, a binary operation on real numbers cannot be a real function because it is of the form $\mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$; that is, the elements of its domain are not single real numbers, but ordered pairs. The most important special feature of real functions as far as their analysis is concerned is that they can be graphed in the Cartesian coordinate plane.

A small number of categories of real functions dominate the study of precalculus and calculus.

i. linear, quadratic, and, more generally, the *polynomial functions*, that is, functions that can be expressed as

$$y = p_n(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \qquad \text{for all } x \text{ in } R,$$

with real coefficients;

ii. *rational functions*, that is, the functions f that can be expressed as

$$f(x) = \frac{p(x)}{q(x)}, \qquad \text{for all } x \text{ in } R \text{ such that } q(x) \neq 0,$$

where $p(x)$ and $q(x)$ are polynomials with real coefficients;